Lightning Talk: Making It Easier To Understand Research Software Impact

Neil P. Chue Hong
Software Sustainability Institute
University of Edinburgh
JCMB, Peter Guthrie Tait Road
Edinburgh, EH9 3FD, United Kingdom
N.ChueHong@software.ac.uk
ORCID: 0000-0002-8876-7606

Abstract— **How can we make it easier to understand research software impact? Is it easy for new researchers to start research in this area? In particular, are the tools available that would let them generate the research that we as the research software community require to convince our funders, fellow researchers and the public at large that having sustainable, open-use research software is important?**

Index Terms — **software, research software, software metrics, software impact.**

I. BACKGROUND

In this lightning talk I will pitch the idea of the WSSSPE community working together to define a conceptual framework and architecture for the tools that are useful to measure research software impact for instance:

- mining scholarly literature for mentions of software
- identifying the linkages between pieces of software
- visualising dependencies between pieces of software or between software and other artefacts

There are already a number of tools that have been developed by others. Depsy[1] is a tool which mines papers to find fulltext mentions of software (currently only those written in Python and R) and analyses GitHub repositories to see where software is being used, thus giving a measure of the impact of a piece of research software or library. The Software Discovery Dashboard[2] aims to search multiple code hosting services, such as Zenodo, Figshare, and GitHub, for scientific software and undertake analysis of it, utilising the Codemeta[3] metadata standards for describing scientific software and building on previous work by Mozilla Science Lab, Github and Figshare on Code as a Research Object[4]. ContentMine[5] provides a platform for machine mining academic papers. ScholarNinja[6] provided a way of scraping software metadata. Lagatto[7] is a tool for tracking analysing article level events in scholarly content which has been used to identify GitHub repositories in scientific papers. Libraries.io[8] provides a way identifying open source library dependencies and discovering alternative software. SciCrunch[9] is developing ways of allocating resource identifiers to software using Research Resource Identifiers (RRIDs)[10] to improve citation and discovery.

What I propose is a working group in WSSSPE to:

- Collectively capture all the useful tools that the community is aware of
- Summarise the use cases for tools in this area
- Define a conceptual framework / architecture
- Identify any areas of functionality which are missing
- Determine what we should work on first
- Run hackathons / sprints to address low-hanging fruit

As researchers, we use evidence and experiences to test our hypotheses and inform our work. However we are lacking the tools to make this easy in the area of research software. By working together, we can ensure that our efforts are aligned.

ACKNOWLEDGMENT

[1] Depsy: http://depsy.org/
[2] Software Discovery Dashboard: https://github.com/mozillascience/software-discovery-dashboard
[3] CodeMeta: https://github.com/codemeta/codemeta
[4] Code as a Research Object: https://science.mozilla.org/projects/codemeta
[5] ContentMine: http://contentmine.org/

[6] ScholarNinja: https://github.com/ScholarNinja/software_metadata
[7] Lagatto: https://github.com/lagotto/lagotto
[8] Libraries.io: https://libraries.io/
[9] SciCrunch: https://scicrunch.org
[10] Resource Identification Initiative: https://www.force11.org/group/resource-identification-initiative
[11] CodeMeta Participants: http://codemeta.github.io/